Common and country-specific dietary patterns in four European cohort studies

Common and Country-Specific Dietary Patterns in Four European Cohort Studies[1]

Helena F. Balder,*†[2] Mikko Virtanen,** Henny A. M. Brants,* Vittorio Krogh,‡ L. Beth Dixon,†††‡ Frans Tan,# Satu Mannisto,** Rino Bellocco,§ Pirjo Pietinen,** Alicja Wolk,§ Franco Berrino,‡ Piet A. Van den Brandt,† Anne M. Hartman‡‡ and R. Alexandra Goldbohm*

**Department of Nutritional Epidemiology, TNO Nutrition and Food Research, Zeist, The Netherlands; †Department of Epidemiology, Maastricht University, Maastricht, The Netherlands; **Department of Nutrition, National Public Health Institute, Helsinki, Finland; ‡Epidemiology Unit, National Cancer Institute, Milan, Italy; ††Department of Nutrition, Food Studies, and Public Health, New York University, New York, NY; ‡‡Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD; #Department of Methodology and Statistics, Maastricht University, Maastricht, The Netherlands; and §Institute of Environmental Epidemiology, Karolinska Institutet, Stockholm, Sweden*

ABSTRACT The association between diet and cancer, predominantly investigated univariately, has often been inconsistent, possibly because of the large number of candidate risk factors and their high intercorrelations. Analysis of dietary patterns is expected to give more insight than analysis of single nutrients or foods. This study aimed to develop and apply a common methodological approach to determine dietary patterns in four cohort studies originating in Finland, the Netherlands, Sweden and Italy. Food items on each of the food frequency questionnaires were aggregated into 51 food groups, defined on the basis of their position in the diet pattern and possible relevance to cancer etiology. Exploratory factor analysis was used to analyze dietary patterns. Using a standardized approach, 3–5 stable dietary patterns were identified, explaining 20–29% of total variance in consumption of the food groups. Two dietary patterns, which explained most of the variance, were consistent across the studies. The first pattern was characterized by high consumption of (salad) vegetables, the second by high consumption of pork, processed meat and potatoes. In addition, a few specifically national food patterns were identified. Sensitivity analyses showed that the identified patterns were robust for number of factors extracted, distribution of input variables and energy adjustment. Our findings suggest that some important eating patterns are shared by the four populations under study, whereas other eating patterns are population specific. J. Nutr. 133: 4246–4251, 2003.

KEY WORDS: • *dietary patterns* • *factor analysis* • *principal components analysis* • *cohort studies* • *cancer*

Epidemiologic and other studies suggest that diet is an important factor in the etiology of different types of cancer. However, the association with various foods and nutrients, predominantly investigated univariately, has often been inconsistent. Apart from the obvious differences in sampled populations, possible explanations for these inconsistent findings could be significant associations by chance due to the large number of candidate foods, inseparable effects due to high intercorrelation among many dietary components or residual confounding or interaction not specified in the model (1–6).

Analysis of dietary patterns is a relatively new approach that allows studying the effect of many foods and their combinations simultaneously. Several publications suggest that analysis of dietary patterns gives more insight into the relation between diet and cancer than analysis of single nutrients or foods; this requires further investigation (7–14). One of the methods that can be used to study dietary patterns is factor analysis. Factor analysis is a multivariate, modeling technique with which to examine underlying patterns in a number of observed variables. The aim can be exploration or confirmation of the structure of associations in a set of variables, or data reduction (i.e., to describe the data with fewer variables). Factor analysis models can result in factors that are, unlike the

[1] Supported by a grant from the Commission of the European Communities (QLK1-CT-1999–00575). NLCS was supported by a grant from the Dutch Cancer Society (TNO V 1999–1963). ATBC was supported by the National Cancer Institute. ORDET was supported by grants from the Italian League against Cancer, by the Italian Ministry of Health and by the Italian Association of Cancer Research (AIRC). SMC was supported by grants from the Swedish Research Council/ Longitudinal Studies, by grants from the Swedish Cancer Society and the World Cancer Research Foundation and the Swedish Foundation for International Cooperation in Research and Higher Education.

[2] To whom correspondence should be addressed.
E-mail: balder@voeding.tno.nl.

Manuscript received 13 June 2003. Initial review completed 23 July 2003. Revision accepted 12 September 2003.

TABLE 1

Characteristics of the four European cohort studies (1985–1992) participating in the DIETSCAN project[1]

	ATBC	NLCS	SMC	ORDET
Country	Finland	the Netherlands	Sweden	Italy
Baseline year	1985–1988	1986	1987–1990	1987–1992
Baseline cohort size, *n*	29,133	120,852[2]	66,651	10,788
Sex	Men	Men and women	Women	Women
Age range, *y*	50–69	55–69	40–74	35–69
FFQ				
Total items, *n*	276	150	67	107
Reference period, *mo*	12	12	6	12
Frequency	Units/times per d/wk/mo	7 categories (never to 6–7 times/wk)	8 categories (never/seldom to >4 times/d)	Times per wk/mo
Quantification	Portion size picture booklet (3–5 per item)	Natural or household units (fixed weight per unit)	Age-specific standard portion sizes	Portion size pictures (less/equal/more than 1–3 pictures)

[1] Abbreviations: ATBC, Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; NLCS, Netherlands Cohort Study on Diet and Cancer; SMC, Swedish Mammography Cohort; ORDET, ORmoni e Dieta nella Eziologia dei Tumori.

[2] 62,573 women and 58,279 men. Because of the case-cohort design, the subjects used in the analyses were derived from a random sample from the cohort.

original variables, uncorrelated, or can allow for correlation among factors.

The authors are engaged in a multicenter project including four ongoing European cohort studies on *diet*ary patterns and *can*cer (DIETSCAN).[3] The purpose of the overall project is to investigate whether dietary pattern analysis reveals patterns that are consistently associated with specific cancers across four populations differing in dietary habits. Such a finding would strongly support the hypothesis that specific food patterns rather than individual foods contribute to the risk of some cancers.

Despite the accumulating number of nutritional epidemiologic studies that have conducted factor analysis, little research has been done to assess the influence of (mostly subjective) decisions taken during the analytic process, such as the grouping of input variables or the number of extracted variables on the resulting patterns. Assessing this influence is of particular importance for the comparison of patterns across studies to be able to attribute any differences to true differences among the populations and not to artifacts resulting from the different instruments and criteria used.

The objective of the present paper was to develop a common approach for the different data sets, to evaluate the influence of arbitrary analytical decisions on the identified patterns and to compare similarities and differences in dietary patterns among four different European populations.

SUBJECTS AND METHODS

Four ongoing European cohort studies on diet and cancer were identified that fulfilled the following criteria: *1*) prospective cohort study investigating the effects of diet on the risk of cancer; *2*) comprehensive diet assessment that included items that cover the entire diet; and *3*) a validated dietary assessment instrument. The Alpha Tocopherol Beta Carotene Cancer Prevention Study (ATBC, Finland), the Netherlands Cohort Study on diet and cancer (NLCS), the Swedish Mammography Cohort (SMC) and the Ormoni e Dieta nella Eziologia dei Tumori (ORDET, Italy) fulfilled these criteria. All four studies were established between 1985 and 1992 with follow-up through record linkage of the cohorts with the national or local cancer registries. Characteristics of the cohort studies and the food frequency questionnaires (FFQ) are summarized in **Table 1**.

The ATBC Cancer Prevention Study, Finland. The ATBC Cancer Prevention Study was a randomized placebo-controlled intervention study conducted among male smokers (15,16). The cohort consisted of 29,133 white men aged 50–69 y at baseline who smoked ≥5 cigarettes/d and were living in southwestern Finland. Subjects were recruited between 1985 and 1988 and the intervention, consisting of supplementation with α-tocopherol and β-carotene, ended after 5–8 y in 1993. The follow-up on cancers and deaths was continued after the intervention period. Before randomization, a questionnaire was administered on food consumption and other risk factors for cancer; 27,111 participants satisfactorily completed a self-administered 276-item FFQ. Subjects were asked to report frequency of consumption over the past 12 mo and to estimate portion sizes using a color picture booklet. This dietary instrument, which has been evaluated for reproducibility and validity (17), is used in the analysis.

The Netherlands Cohort Study, the Netherlands. The NLCS is a prospective cohort study that began in September 1986 (18). The cohort included 62,573 women and 58,279 men aged 55–69 y at the beginning of the study. The study population originated from 204 municipalities with computerized population registries. At baseline, the cohort members completed a mailed, self-administered questionnaire on dietary habits and other risk factors for cancer. The nutrition part of the questionnaire was a semiquantitative FFQ that asked about the habitual consumption of 150 food items in the previous year. The FFQ was validated by 9-d dietary records in a subsample of the cohort (19). Because of the case-cohort design, a random subcohort of 3500 subjects (1688 men, 1812 women) was sampled from the cohort after the baseline exposure measurement. Because of missing or inconsistent dietary data, the analyses are based on data of 3123 subcohort members (1525 men, 1598 women).

The Swedish Mammography Study, Sweden. The SMC originated from population-based mammography screening and started in 1987–1990 in the counties of Uppsala and Västmanland in Sweden. A questionnaire, together with a mailed invitation to be screened by mammography, was sent to all women aged 40–76 y, and was returned by 66,651 (73.8%) women in the source population. The self-administered FFQ asked how often, on average, during the past 6 mo, participants consumed 67 commonly eaten foods. For calculation of daily consumption, age-specific standardized portion sizes were used, based on mean values from 5922 d of weighed food records among 213 women. In a subsample of the cohort (n = 129), the food frequency data were validated. The design and methods are described in more detail elsewhere (20).

[3] Abbreviations used: ATBC, Alpha-Tocopherol Beta-Carotene Cancer Prevention Study; DIETSCAN, DIETary patternS and CANcer; EFA, exploratory factor analysis; FFQ, food frequency questionnaires; NLCS, Netherlands Cohort Study on diet and cancer; ORDET, Ormoni e Dieta nella Eziologia dei Tumori; SMC, Swedish Mammography Cohort.

The ORDET Study, Italy. The ORDET study is a prospective cohort study among Italian women on hormonal factors and diet in the etiology of breast cancer. The study population was recruited in 1987–1992 and included 10,788 healthy volunteers. All participants were female residents of the Province of Varese in Northern Italy, aged 35–69 y. A lifestyle questionnaire included a 107-item semiquantitative FFQ (26 food items quantified by photos, 17 items on fat added to dishes). The validity and reproducibility of this questionnaire was described elsewhere (21,22). The analyses are based on 9208 participants with complete dietary data.

Human subjects. ATBC was approved by the institutional review boards of the National Public Health Institute, Finland and the U.S. National Cancer Institute. NLCS was approved by the institutional review boards of TNO Toxicology and Nutrition Institute and the University of Limburg, the Netherlands. ORDET was approved by the ethical review board of the Italian National Cancer Institute, Italy. SMC was approved by the ethics committee at Uppsala University Hospital and by the regional ethics committee of the Karolinska Institute, Sweden.

Standardization of dietary data and food grouping. The dietary data in the four studies were derived from FFQ with different degrees of detail and different numbers of items. Because the number and detail of food variables used as input in the factor analysis is likely to influence the resulting patterns, a common food grouping was developed. Food items of the four FFQ were aggregated into 51 food groups (Appendix, Table 1),[4] which were defined on the basis of their role in the diet and their possible relevance to cancer etiology. In addition to the standardized food groups, some country-specific foods were included to prevent loss of possibly relevant details. These food groups were an important part of the corresponding cohort's diet, whereas they were relatively unimportant in the other cohorts; therefore, the differences for some items are due not to missing data but to the rare consumption of these items by some cohorts.

Data analyses. All statistical analyses were performed separately for the four cohort studies and for men and women. Because only the NLCS included both men and women, this resulted in five study data sets. Subjects with complete dietary data were used in the analysis. All data sets had a sufficient number of observations according to the criteria published by Hatcher (23), who recommended at least five subjects per input variable in factor analysis.

To identify dietary patterns, exploratory factor analysis (EFA) was performed on the correlation matrix of the 51 food groups. The analyses were performed using the SAS System (SAS Institute, Cary, NC), with the exception of the Italian data, which were analyzed using Stata Software (Stata, College Station, TX). The results of both statistical packages were compared for the ORDET and NLCS data sets and the factor solutions yielded by SAS (Proc Factor procedure) and Stata (factor pcf) were very similar.

To examine the influence of analytic decisions on the stability of the dietary patterns, sensitivity analyses were conducted with the number of factors extracted, dichotomization of extremely skewed variables and energy adjustment of the food group variables by the residual method (24), as explained below.

One of the most important decisions in factor analysis is the choice of the number of factors to be extracted. A factor is usually extracted by default if its eigenvalue (a measure of the amount of variance that is accounted for) is >1. A useful additional criterion is the point at which the scree plot (plot of the eigenvalues against the number of factors) levels off. However, this plot does not always show a clear break and may be subject to sampling variation. To assess the effect of extracting an additional factor on the content and interpretation of the previous factors, an increasing number (i.e., 2–6) of factors were extracted and compared for each data set.

Dietary variables often have a large number of one value (zero); because this affects the correlation between the variables, it can lead to spurious factors. Therefore, it was decided to dichotomize variables with >75% of nonusers (nonusers = 0 vs. users = 1) because more contrast was observed in examining whether these food groups were consumed than the quantity in which they were consumed. Other transformations to enhance normality and linearity (such as log or square root) would hamper interpretation of the factor scores calculated for dietary patterns. Furthermore, because extensive data cleaning in all data sets was carried out, potential outliers were considered not to be the result of incorrect data and were not dropped.

Our primary interest was in dietary patterns based on the relative composition of the diet, and not in those based on the total amount consumed. Therefore, we performed pattern analysis with the food-group variables both unadjusted and adjusted for energy, using the residual method of Willett and Stampfer (25), to assess the influence of energy adjustment.

Stability was assessed by comparing the factor solutions between two random halves of each of the five data sets and by comparing the factor solutions across the sensitivity analyses. A factor was considered stable if the food groups with significant contributions were similar and their factor loadings were comparable in both direction (positive/negative) and magnitude.

The final factor solution was determined for each study separately on the basis of stability and interpretability of the patterns resulting from the factor solutions. The resulting number of extracted factors may therefore be different for each data set.

In summary, the factor loading matrices of *1*) untransformed and dichotomized variables, *2*) unadjusted variables and variables adjusted for energy intake, *3*) extracting an increasing number of factors and *4*) two random halves of the data set were compared both visually and with Procrustes rotation, within each study. With Procrustes rotation, two different factor solutions were compared by rotating the loadings matrix of the first so that it was as similar to the other as possible, while retaining the orthogonality (independence). If comparing a 3-factor solution to a 4-factor solution, for example, the smaller loading matrix was appended with zeroes and then rotated. The resulting transformation matrix can be regarded as the correlation matrix between the factors from different solutions (26).

After orthogonal varimax rotation of the factors, food groups with absolute factor loadings > 0.35 were considered in the interpretation of the factors. The larger the factor loading of a given food group, the greater the correlation of that food group to that dietary pattern. A negative factor loading indicates that food groups were inversely associated with the pattern.

RESULTS

The characteristics of the cohort data sets are presented in **Table 2**. A listing of the 51 food groups used in the pattern analyses and their mean consumption (g/d) is in the Appendix (Table 1). Foods not included in the study-specific FFQ as well as country-specific foods included in study-specific FFQ only (e.g., allium vegetables for the Netherlands, other fruits for Italy, pizza for Italy and Finland, and light beer for Sweden) resulted in several missing entries in that table.

Overall, men consumed more potatoes, bread, beer and spirits than women. Finnish men consumed more dairy products, processed meat, fish, eggs, butter and beer. Among both men and women in the Netherlands, consumption of legumes, cabbages and tea was higher than in the other countries. Among women in Sweden, the consumption of dry cereals was much higher than in the other countries. In Italy, women's consumption of raw leaf vegetables, tomatoes, pasta, oil, full cream cheese, beef and veal, and wine was relatively high. Italy also had a high consumption of white bread, whereas in the other countries, mainly brown bread types were consumed.

Model selection. For all datasets, the scree plot of the eigenvalues of the first 20 factors was generated by the EFA (**Fig. 1**). On the basis of the scree plots, 2–6 dietary patterns were identified, explaining up to 29% of the total variance. Because the scree plots did not indicate a single clear break that could be used as an objective criterion to use in choosing the number of factors for any of the studies, for each data set, the 2-, 3-, 4-, 5- and 6-factor solutions were extracted and compared (**Table 3**). In general, increasing the number of

[4] The Appendix data are available in the online posting of this article at www.nutrition.org.

TABLE 2

Characteristics of the subjects with complete dietary data, included in the dietary pattern analysis in the five datasets[1,2]

	ATBC	NLCS		SMC	ORDET
Sex	Men	Men	Women	Women	Women
n	27,111	1525[3]	1598[3]	61,469	9208
Age, *y*	57.7 ± 5.1	61.4 ± 4.2	61.4 ± 4.3	53.7 ± 9.7	48.6 ± 8.6
BMI, *kg/m*^{2}	26.3 ± 3.8	25.0 ± 2.6	25.1 ± 3.5	24.8 ± 4.4	25.4 ± 4.3
Energy intake, *kJ/d*	11192 ± 3138	9022 ± 2147	7055 ± 1715	5563 ± 1573	7430 ± 2170
Excluding alcohol, *kJ/d*	10670 ± 3091	8597 ± 2133	6892 ± 1699	5473 ± 1567	7137 ± 2121
Smoking, %					
Non	0	13	59	54[4]	64
Ex	0	51	20	27	16
Current	100	36	21	19	20

[1] Values are means ± SEM or %.
[2] Abbreviations; ATBC, Alpha Tocopherol Beta-Carotene Cancer Prevention Study (Finland); NLCS, Netherlands Cohort Study on Diet and Cancer (Netherlands); SMC, Swedish Mammography Cohort (Sweden); ORDET, ORmoni e Dieta nella Eziologia dei Tumori (Italy).
[3] Because of the case-cohort design, the subjects used in the analyses were derived from a random sample from the cohort.
[4] Smoking data for 1997 (not available for baseline 1987–90).

extracted factors did not materially change the previously extracted factors. For example, comparing the 4-factor solution with the 5-factor solution of NLCS men, the Procrustes rotation indicated that the first factor ("Salad vegetables, oil, pasta") in the 4-factor solution and the first extracted factor in the 5-factor solution had a correlation of 0.97. Hence, this factor was interpreted to be the same. Similarly, factor 2 ("Cooked vegetables") and factor 4 ("Pork, processed meat, coffee") were interpreted to remain the same, based on correlations of 0.84 and 0.96, respectively. Occasionally, when an additional factor was extracted, some food groups shifted between the factors. For example the greatest variation of factor 3 ("Sweet foods") came from the old factor 3, but also from 2 and 4 with respective correlations of 0.42 and −0.22. Some of the variation from factor 5 ("Brown/white bread substitution") came from factors 3 and 2, but much of it was newly accounted for (Table 3).

Dichotomizing variables with a high percentage (>75%) of nonusers did not affect the food groups with significant factor loadings, the magnitude of the factor loadings or the explained variance, and thus the order of the extracted patterns (Table 3). This resulted in correlations of 0.98–1.00 on the diagonal of the Procrustes rotation matrix, and low (<0.10) mutual correlations between the factors.

When using the energy-adjusted food groups in the pattern analysis, the factor solutions were mostly comparable with the unadjusted factor solutions (Table 3). As could be expected, mainly the factors with high loadings on energy-contributing food groups changed. By using energy-adjusted food variables, substitution of foods such as brown vs. white bread and low fat vs. medium and full-fat dairy products became more important. The energy adjustment forces substitution patterns because it requires each food group not be correlated with energy. Except for some minor changes, the other factors were unaffected by energy adjustment. For example, for NLCS men, the factors 1, 2, 4 and 5 had correlations of 0.95–0.98 in the Procrustes

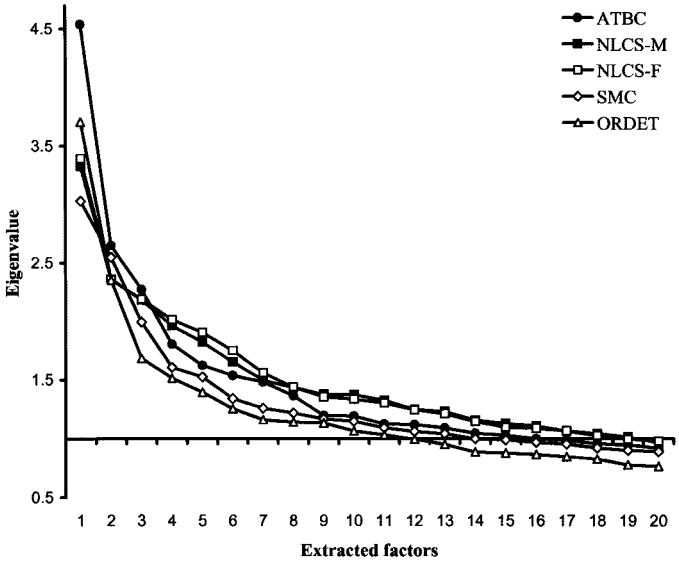

FIGURE 1 Scree plot of the eigenvalues of the first 20 factors identified by exploratory factor analysis, for the four European cohort studies (1985–1992).

TABLE 3

Extraction order and similarity of dietary patterns (determined by Procrustes rotation) identified in exploratory factor analysis, when extracting 2–6 factors and using different sensitivity analyses for 1525 men in the Netherlands Cohort Study on Diet and Cancer (NLCS)

	Number of factors extracted					Transformations		
Factor label	2	3	4	5	6	UU[1]	DU[1,2]	UE[1]
Salad vegetables, oil, pasta	1	1	1	1	1	1	1	1
Cooked vegetables	2	2/3	2	2	2	2	2	2
Pork, processed meat, potatoes		2/3	3/4	4	3	3	3	3
Sweet foods			3/4	3	4	4	4	3
Brown/white bread substitution				5	5	5	5	4
Fat dairy					6	6	6	5
Low/full fat margarine substitution								6

[1] UU: untransformed-unadjusted; DU dichotomized-unadjusted; UE untransformed-energy adjusted.
[2] Dichotomized variables: fermented whole milk and milk products, low fat cheese and cheese spreads, low fat margarine.

rotation matrix. On the basis of these results, we decided to use the variables unadjusted for energy for the final factor solution. The analyses based on split samples were very similar; only the ordering of the factors changed to some extent (results not shown). On the basis of the results of the sensitivity analyses, we determined the final number of extracted factors to be 3 for the ATBC study, 5 for both men and women of the NLCS study, 4 for SMC and 4 for the ORDET study.

Interpretation and comparison. The dietary pattern labels and percentage variance of the final factor solutions in the five data sets are presented in **Table 4**. The factor loadings (>0.35) of the predefined food groups of the identified dietary patterns are in the Appendix (Table 2).

Two factors were qualitatively similar across studies and between men and women, although in some instances specific food groups did not load equally and the amount of variance explained varied. The first comparable factor had high factor loadings on raw leaf vegetables and tomatoes and food groups such as carrots and cabbages and could therefore be interpreted as a "(Salad) Vegetable" pattern. This first pattern also included oil, poultry, rice, pasta and fish, although the factor loadings of these food groups were not consistently >0.35 in all data sets. For NLCS men and ORDET, a separate "Cooked vegetables" pattern was also identified, with high loadings on cooked leaf vegetables, cabbages, legumes and carrots. The second similar, although somewhat less consistent, factor had high loadings on pork, processed meat and potatoes and was hence labeled as such. Other food groups with high loadings in this pattern were eggs, butter and coffee.

In addition to these two comparable factors, other study-specific patterns emerged. For ATBC, SMC and ORDET, an "Alcohol" pattern (high loadings on wine, beer and spirits) was identified. In the NLCS, in both in men and women, a "White/brown bread substitution" pattern was identified. In NLCS women, a "Sweet and/or savory snacks" pattern (savory snacks, nuts, sweets/candies and cakes/cookies) was identified. In NLCS men, this pattern consisted of cakes and cookies, sweet sandwich spread, sweets and candies. In NLCS women, there was also a "Fat dairy" pattern (potatoes, nonfermented whole milk, margarine, sweet sandwich spread) and in SMC a "Margarine/butter substitution" pattern (high positive loading for margarine, negative loading for butter) occurred.

The total percentage variance explained by the extracted factors reflects the number of input variables in the five data sets. When using a smaller number of food groups (e.g., ORDET had 32 food groups vs. NLCS with 49), there is potentially less unique variance than among a larger number of food groups; thus more variance is explained by a similar number of extracted factors.

DISCUSSION

Using a standardized approach, 3–5 stable dietary patterns were identified in cohort studies from Finland, the Netherlands, Sweden and Italy. Two of the identified patterns, i.e., a "(Salad) Vegetable" and a "Pork, processed meat and potatoes" pattern, were relatively consistent across the studies. In addition to these common dietary patterns, some study-specific food patterns were identified. The identified dietary patterns explained 20–29% of the total variance in consumption of the food groups. In contrast to psychometric applications of factor analysis, for example, the items of food consumption questionnaires are not specifically constructed to be highly correlated so as to characterize underlying traits. Other studies that apply factor analysis on dietary data usually find a comparable percentage of total variance explained. Thus, 30% is reasonable for only a few factors. Even more important is the finding that the same or similar factors among the 4 cohorts explained this variance. The importance of these findings will be illustrated through examination of relationships between these factors and sociodemographic factors and specific cancer sites.

The dietary patterns that emerged were easily interpretable and the patterns labeled "(Salad) Vegetable" and "Pork, processed meat and potatoes" comprised a consistent list of the same food groups across the populations. However, the labels assigned are still somewhat arbitrary and represent our interpretation of the data; others may label these dietary patterns differently. Comparing models with a different number of extracted factors proved to be very useful in gaining insight into the data. Studying the dynamics of models helped in restricting the analysis to fewer, meaningful factors.

A potential objection against the use of factor analysis to define dietary patterns is the influence of subjective decisions on the identified eating patterns (27). Differences between the FFQ and differences in the detail and number of input variables were minimized by aggregating the food items into a predefined, common food grouping. To make optimal use of the available dietary data in each study, not all studies performed the EFA on the same number of variables. McCann et al. (28) demonstrated that three methods of reducing detailed dietary data obtained from a single FFQ before pattern analysis did not affect the number or the character of the patterns identified. In our study, the differences between the FFQ were reduced by allocating the original FFQ items to the same food

TABLE 4

Percentage variance explained by the varimax rotated dietary patterns, identified in the exploratory factor analysis in the four European cohort studies (1985–1992)[1]

	(Salad) Vegetables	Pork, processed meat, potatoes	Cooked vegetables	Alcohol	Sweet and/or savory snacks	Brown/white bread substitution	Other
ATBC	9.6	5.3		5.4			
NLCS men	5.6	4.2	4.8		4.3	4.1	
NLCS women	6.3	4.3			3.9	4.3	4.4[2]
SMC	6.9	5.4		5.3			4.2[3]
ORDET	11.6	7.4	4.8	4.7			

[1] Abbreviations: ATBC, Alpha Tocopherol Beta-Carotene Cancer Prevention Study (Finland); NLCS, Netherlands Cohort Study on Diet and Cancer (Netherlands); SMC, Swedish Mammography Cohort (Sweden); ORDET, ORmoni e Dieta nella Eziologia dei Tumori (Italy).

[2] Pattern labeled "Fat dairy."

[3] Pattern labeled "Margarine/butter substitution."

groups, and we showed that the small differences in number of variables and inclusion of country-specific foods did not hamper the comparability of the results.

Furthermore, we observed that several analytic decisions, such as dichotomization and energy adjustment of the input variables and increasing the number of extracted factors, did not materially change the interpretation of the previously extracted factors. Therefore, methodological differences are not likely to hamper comparison with dietary patterns identified in other studies to a large extent.

As in our analysis, other studies observed a vegetable-rich dietary pattern [labeled as "Fruits and vegetables" (29,30); "Salad" (4) or "Prudent" (31)]; and a pattern with high consumption of pork, processed meat and potatoes [labeled as "Western" (31) or "Traditional" (32)].

The other study-specific patterns identified in the present study were also similar to patterns found in previous studies. For example a pattern consisting mainly of "Cooked vegetables" was also observed by Hebert (33) in a U.S. population and by Maskarinec (34) in Hawaiian women. A "Snack" pattern, consisting mainly of sweet foods and/or junk food was observed in several studies (29,30,33,35–37), and an "Alcohol" or "Drinker" pattern was also identified in several studies (30,36,38).

In conclusion, the sensitivity analyses suggest that the dietary pattern approach is robust for energy adjustment, distribution of input variables and number of factors extracted. Our findings suggest that some eating patterns are common to the four European populations under study, but that other eating patterns are country specific. The finding that shared patterns exist across populations living under different social and cultural circumstances, suggests that to some extent, food choices must be driven by endogenous factors (e.g., biological, psychologic), whereas other patterns are more clearly the result of sociocultural circumstances and food availability.

LITERATURE CITED

1. Farchi, G., Mariotti, S., Menotti, A., Seccareccia, F., Torsello, S. & Fidanza, F. (1989) Diet and 20-y mortality in two rural population groups of middle-aged men in Italy. Am. J. Clin. Nutr. 50: 1095–1103.

2. Randall, E., Marshall, J., Graham, S. & Brasure, J. (1989) Frequency of food use data and the multidimensionality of diet. J. Am. Diet Assoc. 89: 1070–1075.

3. Nestle, M. (1996) Fruits and vegetables: protective or just fellow travelers? Nutr. Rev. 54: 255–257.

4. Randall, E., Marshall, J. R., Graham, S. & Brasure, J. (1990) Patterns in food use and their associations with nutrient intakes. Am. J. Clin. Nutr. 52: 739–745.

5. Wirfalt, A.K.E. & Jeffery, R. W. (1997) Using cluster analysis to examine dietary patterns: nutrient intakes, gender, and weight status differ across food pattern clusters. J. Am. Diet Assoc. 97: 272–279.

6. Hu, F. B. (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. Curr. Opin. Lipidol. 13: 3–9.

7. Randall, E., Marshall, J. R., Brasure, J. & Graham, S. (1992) Dietary patterns and colon cancer in western New York. Nutr. Cancer 18: 265–276.

8. Slattery, M. L. & Boucher, K. M. (1998) The senior authors' response: factor analysis as a tool for evaluating eating patterns. Am. J. Epidemiol. 148: 20–21.

9. Palli, D., Russo, A. & Decarli, A. (2001) Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. Cancer Causes Control 12: 163–172.

10. Terry, P., Suzuki, R., Hu, F. B. & Wolk, A. (2001) A prospective study of major dietary patterns and the risk of breast cancer. Cancer Epidemiol. Biomark. Prev. 10: 1281–1285.

11. Terry, P., Hu, F. B., Hansen, H. & Wolk, A. (2001) Prospective study of major dietary patterns and colorectal cancer risk in women. Am. J. Epidemiol. 154: 1143–1149.

12. Fung, T., Hu, F. B., Fuchs, C., Giovannucci, E., Hunter, D. J., Stampfer, M. J., Colditz, G. A. & Willett, W. C. (2003) Major dietary patterns and the risk of colorectal cancer in women. Arch. Intern. Med. 163: 309–314.

13. Handa, K. & Kreiger, N. (2002) Diet patterns and the risk of renal cell carcinoma. Public Health Nutr. 5: 757–767.

14. Masaki, M., Sugimori, H., Nakamura, K. & Tadera, M. (2003) Dietary patterns and stomach cancer among middle-aged male workers in Tokyo. Asian Pac. J. Cancer Prev. 4: 61–66.

15. The ATBC Cancer Prevention Study Group (1994) The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. Ann. Epidemiol. 4: 1–10.

16. The Alpha-Tocopherol Beta-Carotene Cancer Prevention Study Group (1994) The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. N. Engl. J. Med. 330: 1029–1035.

17. Pietinen, P., Hartman, A. M., Haapa, E., Rasanen, L., Haapakoski, J., Palmgren, J., Albanes, D., Virtamo, J. & Huttunen, J. K. (1988) Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. Am. J. Epidemiol. 128: 655–666.

18. Van den Brandt, P. A., Goldbohm, R. A., Van't Veer, P., Volovics, A., Hermus, R. J. & Sturmans, F. (1990) A large-scale prospective cohort study on diet and cancer in The Netherlands. J. Clin. Epidemiol. 43: 285–295.

19. Goldbohm, R. A., Van den Brandt, P. A., Brants, H. A., Van't Veer, P., Al, M., Sturmans, F. & Hermus, R. J. (1994) Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur. J. Clin. Nutr. 48: 253–265.

20. Wolk, A., Bergstrom, R., Hunter, D., Willett, W., Ljung, H., Holmberg, L., Bergkvist, L., Bruce, A. & Adami, H. O. (1998) A prospective study of association of monounsaturated fat and other types of fat with risk of breast cancer. Arch. Intern. Med. 158: 41–45.

21. Faggiano, F., Vineis, P., Cravanzola, D., Pisani, P., Xompero, G., Riboli, E. & Kaaks, R. (1992) Validation of a method for the estimation of food portion size. Epidemiology 3: 379–382.

22. Pisani, P., Faggiano, F., Krogh, V., Palli, D., Vineis, P. & Berrino, F. (1997) Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. Int. J. Epidemiol. 26 (suppl. 1): S152–S160.

23. Hatcher, L. (1994) A Step-By-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling. SAS Institute, Cary, NC.

24. Willett, W. (1990) Food frequency methods. In: Nutritional Epidemiology (Willett, W., ed.). Oxford University Press, Oxford, UK.

25. Willett, W. C. & Stampfer, M. J. (1986) Total energy intake: implications for epidemiologic analyses. Am. J. Epidemiol. 124: 17–27.

26. Mardia, K. V., Kent, J. T. & Bibby, M. C. (1979) Multivariate Analysis. Academic Press, London, UK.

27. Martinez, M. E., Marshall, J. R. & Sechrest, L. (1998) Invited commentary: factor analysis and the search for objectivity. Am. J. Epidemiol. 148: 17–19.

28. McCann, S. E., Marshall, J. R., Brasure, J. R., Graham, S. & Freudenheim, J. L. (2001) Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutr. 4: 989–997.

29. Nicklas, T. A., Webber, L. S., Thompson, B. & Berenson, G. S. (1989) A multivariate model for assessing eating patterns and their relationship to cardiovascular risk factors: the Bogalusa Heart Study. Am. J. Clin. Nutr. 49: 1320–1327.

30. Schulze, M. B., Hoffmann, K., Kroke, A. & Boeing, H. (2001) Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. Br. J. Nutr. 85: 363–373.

31. Hu, F. B., Rimm, E., Smith-Warner, S. A., Feskanich, D., Stampfer, M. J., Ascherio, A., Sampson, L. & Willett, W. C. (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am. J. Clin. Nutr. 69: 243–249.

32. Barker, M. E., McClean, S. I., Thompson, K. A. & Reid, N. G. (1990) Dietary behaviours and sociocultural demographics in Northern Ireland. Br. J. Nutr. 64: 319–329.

33. Hebert, J. R., Stoddard, A. M., Harris, D. R., Sorensen, G., Hunt, M. K., Morris, D. H. & Ockene, J. K. (1993) Measuring the effect of a worksite-based nutrition intervention on food consumption. Ann. Epidemiol. 3: 629–635.

34. Maskarinec, G., Novotny, R. & Tasaki, K. (2000) Dietary patterns are associated with body mass index in multiethnic women. J. Nutr. 130: 3068–3072.

35. French, S. A., Perry, C. L., Leon, G. R. & Fulkerson, J. A. (1994) Food preferences, eating patterns, and physical activity among adolescents: correlates of eating disorders symptoms. J. Adolesc. Health 15: 286–294.

36. Matalas, A. L., Franti, C. E. & Grivetti, L. E. (1999) Comparative study of diet and disease prevalence in Greek Chians. Part I: Rural and urban residents of Chios. Ecol. Food Nutr. 38: 351–380.

37. North, K. & Emmett, P. (2000) Multivariate analysis of diet among three-year-old children and associations with socio-demographic characteristics. The Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC) Study Team. Eur. J. Clin. Nutr. 54: 73–80.

38. Slattery, M. L., Boucher, K. M., Caan, B. J., Potter, J. D. & Ma, K. N. (1998) Eating patterns and risk of colon cancer. Am. J. Epidemiol. 148: 4–16.